

REPORT DOCUMENTATION PAGE			Form Approved OMB NO. 0704-0188	
<small>Public reporting burden for this collection of information is estimated to average 1 hour per response, including the time for reviewing instructions, searching existing data sources, gathering and maintaining the data needed, and completing and reviewing the collection of information. Send comment regarding this burden estimate or any other aspect of this collection of information, including suggestions for reducing this burden, to Washington Headquarters Services, Directorate for Information Operations and Reports, 1215 Jefferson Davis Highway, Suite 1204, Arlington, VA 22202-4302, and to the Office of Management and Budget, Paperwork Reduction Project (0704-0188), Washington, DC 20503.</small>				
1. AGENCY USE ONLY (Leave blank)		2. REPORT DATE February 1996		3. REPORT TYPE AND DATES COVERED <i>Technical Report</i>
4. TITLE AND SUBTITLE Bootstrap by Sequential Resampling			5. FUNDING NUMBERS <i>DAAH04-96-1-0082</i>	
6. AUTHOR(S) C.R. Rao, P.K. Pathak and V.I. Koltchinskii				
7. PERFORMING ORGANIZATION NAME(S) AND ADDRESS(ES) Center for Multiivariate Analysis 417 Classroom Building Penn State University University Park, PA 16802				
9. SPONSORING / MONITORING AGENCY NAME(S) AND ADDRESS(ES) U.S. Army Research Office P.O. Box 12211 Research Triangle Park, NC 27709-2211			10. SPONSORING / MONITORING AGENCY REPORT NUMBER <i>ARO 30529.87-MA-SOI</i>	
11. SUPPLEMENTARY NOTES The views, opinions and/or findings contained in this report are those of the author(s) and should not be construed as an official Department of the Army position, policy or decision, unless so designated by other documentation.				
12a. DISTRIBUTION / AVAILABILITY STATEMENT Approved for public release; distribution unlimited. This paper examines methods of resampling for bootstrap from a survey sampling point of view. Given an observed sample of size n , resampling for bootstrap involves n repeated trials of simple random sampling with replacement. From the point of view of information content, it is well known that simple random sampling with replacement does not result in samples that are equally informative (see Pathak (1964)). This is due to different numbers of distinct observations occurring in different bootstrap samples. We propose an alternative scheme of sampling sequentially (with replacement each time) until k distinct original observations appear. In such a scheme, the bootstrap sample size becomes random as it varies from sample to sample, but each sample has exactly the same number of distinct observations. We show that the choice of $k = (1 - e^{-1})n \sim .632n$ has some advantage, stemming from the observation made by Efron (1983) that the usual bootstrap samples are supported on approximately $.632n$ of the original data points. Using recent results on empirical processes, we show that main empirical characteristics of the sequential resampling bootstrap are asymptotically within the distance of order $\sim n^{-3/4}$ of the corresponding characteristics of the usual bootstrap.			12 b. DISTRIBUTION CODE	
14. SUBJECT TERMS bootstrap, sequential resampling, information content, sampling viewpoint, asymptotic correctness, empirical measures, empirical processes, weak convergence, symmetrization, Poissonization, p-variation, quantiles			15. NUMBER OF PAGES 26	
17. SECURITY CLASSIFICATION OR REPORT UNCLASSIFIED			16. PRICE CODE	
18. SECURITY CLASSIFICATION OF THIS PAGE UNCLASSIFIED		19. SECURITY CLASSIFICATION OF ABSTRACT UNCLASSIFIED		20. LIMITATION OF ABSTRACT UL

BOOTSTRAP BY SEQUENTIAL RESAMPLING

C.R. Rao, P.K. Pathak and V.I. Koltchinskii

Technical Report 96-01

February 1996

Center for Multivariate Analysis
417 Classroom Building
Penn State University
University Park, PA 16802

Research sponsored by the Army Research Office under Grant DAAHO4-96-1-0082. The United States Government is authorized to reproduce and distribute reprints for governmental purposes notwithstanding any copyright notation hereon.

BOOTSTRAP BY SEQUENTIAL RESAMPLING

C.Radhakrishna Rao
The Pennsylvania State University
University Park, PA 16802

P.K. Pathak and V.I. Koltchinskii
University of New Mexico
Albuquerque, NM 87131

ABSTRACT

This paper examines methods of resampling for bootstrap from a survey sampling point of view. Given an observed sample of size n , resampling for bootstrap involves n repeated trials of simple random sampling with replacement. From the point of view of information content, it is well known that simple random sampling with replacement does not result in samples that are equally informative (see Pathak (1964)). This is due to different numbers of distinct observations occurring in different bootstrap samples. We propose an alternative scheme of sampling sequentially (with replacement each time) until k distinct original observations appear. In such a scheme, the bootstrap sample size becomes random as it varies from sample to sample, but each sample has exactly the same number of distinct observations. We show that the choice of $k = (1 - e^{-1})n \sim .632n$ has some advantage, stemming from the observation made by Efron (1983) that the usual bootstrap samples are supported on approximately $.632n$ of the original data points. Using recent results on empirical processes, we show that main empirical characteristics of the sequential resampling bootstrap are asymptotically within the distance of order $\sim n^{-3/4}$ of the corresponding characteristics of the usual bootstrap.

Keywords and phrases: bootstrap, sequential resampling, information content, sampling viewpoint, asymptotic correctness, empirical measures, empirical processes, weak convergence, symmetrization, Poissonization, p -variation, quantiles

AMS 1991 subject classification: Primary 62G09; Secondary 62L10

1. INTRODUCTION

Let $\mathcal{S} = (X_1, X_2, \dots, X_n)$ be a random sample from a distribution F , and $\theta(F)$ be a given parameter of interest. Let F_n denote the empirical distribution function based on \mathcal{S} , and suppose that $\theta(F_n)$ is to be used as an estimator of $\theta(F)$. Then, for a large class of functions θ , Efron's bootstrap resampling method (1979, 1982) provides a robust method of evaluating the performance characteristics of $\theta(F_n)$, solely on the basis of information derived (through randomization) from the observed sample \mathcal{S} . As an illustrative example, consider the simple case of

$$\theta(F) = \int x dF = \mu(F), \quad \text{say.} \quad (1.1)$$

Let

$$\sigma^2(F) = \int (x - \mu(F))^2 dF, \quad (1.2)$$

$$\mu_n = \frac{1}{n} \sum X_i \quad (1.3)$$

and

$$\sigma_n^2 = \frac{1}{n} \sum (X_i - \mu_n)^2 \quad (1.4)$$

Then $\mu_n = \mu(F_n)$ and $\sigma_n^2 = \sigma^2(F_n)$. Let

$$\Pi_n = \sqrt{n}(\mu_n - \mu)/\sigma_n \quad (1.5)$$

$$= (\mu(F_n) - \mu(F))/\sqrt{\sigma^2(F_n)} \quad (1.6)$$

The central limit theorem entails that the sampling distribution of Π_n can be approximated by the standard normal distribution. On the other hand, the bootstrap method furnishes an alternative approach to estimating the sampling distribution of Π_n (by repeated resampling of the observed sample \mathcal{S}) as follows:

Given the observed sample \mathcal{S} , select a simple random sample with replacement (SRSWR) of size n from the observed sample \mathcal{S} . Let $\hat{\mathcal{S}}_n = (\hat{X}_1, \dots, \hat{X}_n)$ denote a sample drawn from \mathcal{S} in this fashion. Let \hat{F}_n be the empirical distribution based on $\hat{\mathcal{S}}_n$. Let

$$\begin{aligned} \hat{\Pi}_n &= \sqrt{n}(\mu(\hat{F}_n) - \mu(F_n))/\sqrt{\sigma^2(\hat{F}_n)} \\ &= \sqrt{n}(\hat{\mu}_n - \mu_n)/\hat{\sigma}_n, \quad \text{say.} \end{aligned} \quad (1.7)$$

Then for large n , the conditional distribution of $\hat{\Pi}_n$ given \mathcal{S} is close to that of Π_n . Thus, one can replace the sampling distribution of Π_n by the conditional distribution of $\hat{\Pi}_n$ given \mathcal{S} . In practice, the conditional distribution of $\hat{\Pi}_n$ is estimated by the frequency distribution

(ensemble) of $\hat{\Pi}_n$ obtained by repeated resamplings of \mathcal{S} by SRSWR a large number of times; the observed sampling distribution of $\hat{\Pi}_n$ is referred to as the bootstrap distribution of Π_n . Thus for example, for large n , the quantiles from the bootstrap distribution of Π_n can be used to set up a confidence interval for μ based on the pivot Π_n .

It is well known that owing to the with replacement nature of SRSWR, not all of the observations in bootstrap samples $\hat{\mathcal{S}}_n$, say, will be based on distinct observations from \mathcal{S} . In fact, the information content of $\hat{\mathcal{S}}_n$, the set of distinct observations in $\hat{\mathcal{S}}_n$, is a random variable. Let ν_n denote the number of distinct observations in $\hat{\mathcal{S}}_n$. Then it is easily seen that

$$E(\nu_n) = n[1 - (1 - \frac{1}{n})^n] \quad (1.8)$$

$$\begin{aligned} &\simeq n[1 - e^{-1}] \\ &\simeq n(.632). \end{aligned} \quad (1.9)$$

$$\begin{aligned} \text{Var}(\nu_n) &= \text{Var}(n - \nu_n) \\ &= n(1 - \frac{1}{n})^n + n(n-1)(1 - \frac{2}{n})^n - n^2(1 - \frac{1}{n})^{2n} \end{aligned} \quad (1.10)$$

$$\begin{aligned} &\simeq ne^{-1} + n(n-1)e^{-2} - n^2e^{-2} \\ &\simeq ne^{-1}(1 - e^{-1}) \\ &\simeq n(.368)(.632) \simeq n(.233) \end{aligned} \quad (1.11)$$

so that

$$\sigma(\nu_n) \simeq .482\sqrt{n}. \quad (1.12)$$

In fact, ν_n approaches a binomial distribution $b(n, p)$ with $p \simeq .632$. This shows that in a bootstrap sample, approximately sixty percent of the information from the observed sample \mathcal{S} is utilized while the rest of the forty percent of the data in it is simply randomization. Besides, the 2σ -limits for ν_n are approximately $(.632)n \pm 2(.482)\sqrt{n} \simeq (.632)n \pm \sqrt{n}$. Therefore in approximately 95% of resamplings, the essential supremum of ν_n can be as large as $(.632)n + \sqrt{n}$, and the essential infimum as small as $(.632)n - \sqrt{n}$. Thus for $n=100$, this means that one can expect ν_n to roughly vary in the range (53, 73). Viewed from a purely information content of the bootstrap samples, this variability is neither necessary nor desirable. As an alternative to a fixed size bootstrap samples, we examine in what follows the following sequential approach.

2. A SEQUENTIAL RESAMPLING APPROACH

To select a bootstrap sample, draw observations from \mathcal{S} sequentially by SRSWR until there are $(m+1) \approx n(1 - e^{-1}) + 1$ distinct observations in the observed bootstrap sample, the last observation in the bootstrap sample is discarded to ensure simplicity in technical details. Thus an observed bootstrap sample has the form:

$$\hat{\mathcal{S}}_N = (\hat{X}_1, \hat{X}_2, \dots, \hat{X}_N) \quad (2.1)$$

in which $\hat{X}_1, \dots, \hat{X}_N$ has $m \approx n(1 - e^{-1})$ distinct observations from \mathcal{S} .

The number of distinct observations in $\hat{\mathcal{S}}_N$ is precisely $[n(1 - e^{-1})]$; it is no longer a random variable. It is worth noting that the size N of the bootstrap sample is a random variable with $E(N) \approx n$. The pivot based on $\hat{\mathcal{S}}_N$ is

$$\hat{\Pi}_N = \sqrt{N}(\mu(\hat{F}_N) - \mu(F_n)) \quad (2.2)$$

in which \hat{F}_N denotes the empirical distribution based on the bootstrap sample $\hat{\mathcal{S}}_N$. Again for simplicity in the exposition, we assume that $\sigma^2(F) = 1$. Now for a comparative study of the pivot $\hat{\Pi}_N$, based on the sequential resampling approach, versus $\hat{\Pi}_n$, based on a bootstrap sample of fixed size n , it is necessary to estimate the order of magnitude of the random variable N . It is easily seen that N admits the following decomposition in terms of independent (geometric) random variables:

$$N = I_1 + I_2 + \dots + I_m \quad (2.3)$$

in which $m = [n(1 - e^{-1})]$ using the notation $[\cdot]$ to indicate the largest integer; $I_1 = 1$, and for each k , $2 \leq k \leq m$,

$$P(I_k = j) = (1 - \frac{k-1}{n})(\frac{k-1}{n})^{j-1} \quad (2.4)$$

for $j = 1, 2, \dots$. Therefore

$$E(N) = n[\frac{1}{n} + \frac{1}{(n-1)} + \dots + \frac{1}{(n-m+1)}]. \quad (2.5)$$

It can be seen that

$$\begin{aligned} E(N) &= n \log \frac{n}{(n-m)} + O(1) \\ &= n + O(1) \end{aligned} \quad (2.6)$$

since $m = [n(1 - e^{-1})]$. Similarly, it is easily seen that

$$\text{Var}(N) = \sum_{k=1}^m \frac{n(k-1)}{(n-k+1)^2} \quad (2.7)$$

$$= n(e-1) + O(1). \quad (2.8)$$

Thus

$$\frac{E(N-n)^2}{n^2} = \frac{(e-1)}{n} + O(\frac{1}{n^2}) \quad (2.9)$$

showing that $(N/n) \rightarrow 1$ in probability. Further analysis similar to that of Mitra and Pathak (Mitra, S.K. and Pathak, P.K. [1984]) can be used to show that

$$\frac{E(\hat{\Pi}_N - \hat{\Pi}_n)^2}{\text{Var}(\Pi_n)} \leq K \sqrt{\frac{\text{Var}(N)}{n^2}} \quad (2.10)$$

$$= O(\frac{1}{\sqrt{n}}). \quad (2.11)$$

This implies that $\hat{\Pi}_n$ and $\hat{\Pi}_N$ are asymptotically equivalent. The preceding computations show that the sequential resampling plan introduced here, in addition to keeping the information content of bootstrap samples constant, also preserves its asymptotic correctness. With this heuristic introduction to the sequential resampling scheme, we turn now to the question of mathematical justification of its validity with the main goal of showing that the sequential bootstrap and the usual one are at a distance $O(n^{-3/4})$, which is enough to justify its asymptotic first order correctness.

Let Γ_2 be the set of probability distributions with finite second moment. We say that distributions $\{F_n\}$ converge in M -sense to F if and only if (a) F_n converges weakly to F and (b) $\int x^2 dF_n$ converges to $\int x^2 dF$. It is easily seen that this notion of convergence is induced by the following M -metric³ on Γ_2 . For each F, G in Γ_2 , define the squared distance $d^2(F, G)$ between F and G as follows:

$$d^2(F, G) = \inf_{X \sim F, Y \sim G} E(X - Y)^2 \quad (2.12)$$

in which the infimum is taken over all pairs of random variables (X, Y) with given marginals F and G respectively (Mallows (1972)).

Now let Y_1, \dots, Y_m be independent random variables with a common distribution G . Let $G^{(m)}$ be the distribution of the standardized variable:

$$s_m = \sqrt{m} \left\{ \frac{1}{m} \sum_{j=1}^m (Y_j - E(Y_j)) \right\}. \quad (2.13)$$

Then the following result holds:

Lemma 2.1 (Mallows).

$$d(G^{(m)}, H^{(m)}) \leq d(G, H). \quad (2.14)$$

The proof of the validity of Efron's bootstrap based on resampling schemes of a preassigned size m can now be seen to be a consequence of the following triangle inequality:

$$d(F_n^{(m)}, \Phi) \leq d(F_n^{(m)}, F^{(m)}) + d(F^{(m)}, \Phi) \quad (2.15)$$

$$\leq d(F_n, F) + d(F^{(m)}, \Phi). \quad (2.16)$$

(Note that $F_n^{(m)}$ here stands for the conditional distribution of the pivot $\hat{\Pi}_m$ (eq.(1.7)), and $F^{(m)}$ stands for the distribution of Π_m based on a sample of size m (eq.(1.5)).)

In this inequality the convergence of $d(F^{(m)}, \Phi)$ to zero essentially follows from the central limit theorem; the convergence of $d(F_n, F)$ to zero is essentially a consequence of the fact that the sample mean converges in mean to the population mean (i.e. $\lim E(\int x dF_n - \int x dF)^2 = 0$). Despite the apparent simplicity of this proof, an added complication that arises in sequential resampling scheme is that the bootstrap sample size " m " is now a random variables; it is denoted by N . Lemma 2.1 is no longer directly applicable. Nevertheless the following extensions of Lemma 2.1 hold.

³known also in the literature as a Wasserstein metric; see, e.g. Rachev (1984)

Lemma 2.2.

$$E\{(\sqrt{N}(\hat{\mu}_N - \mu_n) - \sqrt{n}(\hat{\mu}_N - \mu_n))^2 | \mathcal{S}\} \leq K_1 \frac{s^2}{n} \quad (2.17)$$

where s^2 is the sample variance based on the initial sample and K_1 is a universal constant.

Lemma 2.3

$$E[n(\hat{\mu}_N - \hat{\mu}_n)^2 | \mathcal{S}] \leq K_2 \frac{s^2}{\sqrt{n}}. \quad (2.18)$$

The proofs of these lemmas are based on techniques similar to those of Pathak (1964) and Mitra and Pathak (1984) (cf Lemma 3.1 in Mitra and Pathak).

Lemmas 2.2 and 2.3 imply that the Mallows distance between $\sqrt{N}\hat{\mu}_N$ and $\sqrt{n}\hat{\mu}_n$ converges to zero. This then implies the asymptotic correctness of $\sqrt{N}\hat{\mu}_N$, and thus concludes the heuristic justification for the correctness of the sample sum for the sequential bootstrap.

In the next section we examine the mathematical justification for the preceding heuristic discussion in some detail.

3. MATHEMATICAL JUSTIFICATION: BOOTSTRAPPING EMPIRICAL MEASURES WITH A RANDOM SAMPLE SIZE

Let (X, \mathcal{A}, P) be a probability space, and let (Ω, Σ, γ) be the product of a countable number of copies of (X, \mathcal{A}, P) and the unit interval with the Lebesgue measure. In what follows E denotes the expectation with respect to γ . Consider a sequence $\{X_n : n \geq 1\}$ of random elements defined on (Ω, Σ, γ) as follows:

$$X_n(\omega) = x_n \quad (3.1)$$

in which $\omega = (x_1, x_2, \dots, u) \in \Omega$. Then $\{X_n : n \geq 1\}$ is a sequence of independent random elements with a common distribution P . Let P_n denote the empirical measure based on the sample (X_1, \dots, X_n) from P , i.e.,

$$P_n := n^{-1} \sum_{1 \leq i \leq n} \delta_{X_i} \quad (3.2)$$

with δ_x being the unit point mass at $x \in X$.

Now let $\{\hat{X}_{n,j} : j \geq 1\}$ be a sequence of independent random elements with common distribution P_n ; assume that this sequence is defined on $(\hat{\Omega}, \hat{\Sigma}, \hat{\gamma})$ with \hat{E} denoting the expectation with respect to $\hat{\gamma}$. We refer to this sequence as a sequence of bootstrapped observations, or just a bootstrapped sequence. Given a number $N \geq 1$, let $\hat{P}_{n,N}$ denote the empirical measure based on the bootstrap sample $(\hat{X}_{n,1}, \dots, \hat{X}_{n,N})$ of size N ; i.e.,

$$\hat{P}_{n,N} := N^{-1} \sum_{1 \leq i \leq N} \delta_{\hat{X}_{n,i}}. \quad (3.3)$$

We refer to it as a bootstrapped empirical measure of size N .

In what follows we assume that $N = N_n$ is a random variable defined on the space $(\Omega, \Sigma, \gamma) \times (\hat{\Omega}, \hat{\Sigma}, \hat{\gamma})$ and takes only positive integer values. Thus we allow the size of the bootstrapped sample to be a random variable. The main object of this section is to show that if (N/n) converges to one in probability, then the bootstrapped empirical measure $\hat{P}_{n,N}$ is at a distance of $o(n^{-1/2})$ from the bootstrapped empirical measure $\hat{P}_n := \hat{P}_{n,n}$, so that all of the \sqrt{n} -asymptotic results for the classical bootstrap carry over to the sequential bootstrap with a random sample size. To do so, we need to introduce a few preliminary results on limit theorems for general empirical processes (see Dudley (1984), Giné and Zinn (1986 and 1994) and Pollard (1984)).

A general empirical process is defined as a random signed measure

$$Z_n := \sqrt{n}(P_n - P), \quad n \geq 1, \quad (3.4)$$

on the measurable space (X, \mathcal{A}) . Given a measurable function f and a signed measure λ on (X, \mathcal{A}) , let $\lambda(f)$ denote the integral $\int f d\lambda$, provided $\int |f| d|\lambda| < \infty$. It is convenient to consider the empirical process Z_n as a random function $f \rightarrow Z_n(f)$ defined on a class $\mathcal{F} \subset L_2(X; dP)$. The finite dimensional distributions of the sequence of random functions Z_n converge to the finite dimensional distributions of a Gaussian random function $f \rightarrow G_P(f)$ with mean zero and covariance function

$$D(G_P(f), G_P(g)) = P(fg) - P(f)P(g), \quad f, g \in L_2(X; dP). \quad (3.5)$$

Such a Gaussian random function is often called a P -Brownian bridge. We endow $L_2(X; dP)$ with a metric ρ_P defined by

$$\begin{aligned} \rho_P^2(f, g) &:= D(G_P(f) - G_P(g), G_P(f) - G_P(g)) \\ &= P(f - g)^2 - (P(f - g))^2. \end{aligned} \quad (3.6)$$

Convergence of the sequence $\{Z_n : n \geq 1\}$ to the limit G_P (now in functional sense) has been studied (see, for example, Pollard (1984)) in the space $\ell^\infty(\mathcal{F})$ of all uniformly bounded functions on the class \mathcal{F} with the norm

$$\|Y\|_{\mathcal{F}} := \sup\{|Y(f)| : f \in \mathcal{F}\}, \quad Y \in \ell^\infty(\mathcal{F}). \quad (3.7)$$

Z_n belongs to this space a.s. under the condition

$$\int_X \|\delta_x\|_{\mathcal{F}} P(dx) = \int_X \sup_{f \in \mathcal{F}} |f(x)| P(dx) < +\infty. \quad (3.8)$$

Let $C_{bu}(\mathcal{F})$ be the subspace of all ρ_P -uniformly continuous and uniformly bounded functions on \mathcal{F} .

In the sequel, we use weak convergence of random functions in the space $\ell^\infty(\mathcal{F})$ in the Hoffman-Jørgensen sense. Let ζ_n be a sequence of functions from $\Omega \times \mathcal{F}$ into R^1 , such that for all $n \geq 1$

$$\gamma^*(\{\omega : \|\zeta_n\|_{\mathcal{F}} = +\infty\}) = 0 \quad (3.9)$$

in which γ^* denotes the outer measure induced by γ (note the need for γ^* since ζ_n is not necessarily a random element (measurable)). Then the sequence ζ_n is said to converge weakly in $\ell^\infty(\mathcal{F})$ to a limit $\zeta : \Omega \times \mathcal{F} \mapsto R^1$ iff

- (a) $\gamma^*(\{\zeta \notin C_{bu}(\mathcal{F})\}) = 0$;
- (b) for any bounded and $\|\cdot\|_{\mathcal{F}}$ -continuous functional $\Phi : \ell^\infty(\mathcal{F}) \mapsto R^1$, we have

$$E^*\Phi(\zeta_n) \rightarrow E\Phi(\zeta), \text{ as } n \rightarrow \infty$$

in which E^* is the outer expectation induced by the expectation operator E . The main results of the classical theory of weak convergence of continuous processes hold for this extension as well; of particular interest is the following result: Let $BL_1(\ell^\infty(\mathcal{F}))$ be the set of all functionals $\Phi : \ell^\infty(\mathcal{F}) \mapsto R^1$ such that

$$\|\Phi\|_\infty := \sup_{y \in \ell^\infty(\mathcal{F})} |\Phi(y)| \leq 1$$

and for all $y_1, y_2 \in \ell^\infty(\mathcal{F})$

$$|\Phi(y_1) - \Phi(y_2)| \leq \|y_1 - y_2\|_{\mathcal{F}}.$$

Given two functions $\zeta_1, \zeta_2 : \Omega \times \mathcal{F} \mapsto R^1$, define the following distance between ζ_1 and ζ_2 :

$$d_\gamma(\zeta_1, \zeta_2) := \sup_{\Phi \in BL_1(\ell^\infty)} |E^*\Phi(\zeta_1) - E^*\Phi(\zeta_2)|. \quad (3.10)$$

Then a sequence $\{\zeta_n : n \geq 1\}$ converges weakly in the space $\ell^\infty(\mathcal{F})$ to the limit ζ if and only if

$$\gamma^*(\{\zeta \notin C_{bu}(\mathcal{F})\}) = 0$$

and

$$\lim_{n \rightarrow \infty} d_\gamma(\zeta_n, \zeta) = 0.$$

A class $\mathcal{F} \subset L_2(X; dP)$ is called P -pregaussian iff there is a version of P -Brownian bridge G_P such that

$$\gamma^*(\{G_P \notin C_{bu}(\mathcal{F})\}) = 0.$$

This class is called P -Donsker iff the sequence of empirical processes Z_n converges weakly in the space $\ell^\infty(\mathcal{F})$ to the P -Brownian bridge G_P . In this case we say that $\mathcal{F} \in CLT(P)$. The following result is well known (Giné and Zinn (1986)). If $\mathcal{F} \in CLT(P)$, then

$$E^* \left\| \sum_{1 \leq i \leq n} (\delta_{X_i} - P) \right\|_{\mathcal{F}} = O(n^{1/2}). \quad (3.11)$$

It is also clear that $\mathcal{F} \in CLT(P)$ iff \mathcal{F} is P -pregaussian and

$$d_\gamma(Z_n, G_P) \rightarrow 0 \text{ as } n \rightarrow \infty.$$

In what follows, we use several well known inequalities for sums of independent random elements in Banach spaces. We assume that all random variables and elements in the following three lemmas are defined on a certain probability space (Ω_Q, Σ_Q, Q) . We denote

the corresponding expectation operator E_Q . Let B be a Banach space. We start with the following symmetrization lemma (Giné and Zinn (1986)).

Lemma 3.1 (Symmetrization lemma): Let $\{Y_n : n \geq 1\}$ be a sequence of independent and identically distributed random elements in B with finite first moment. Let $\{\varepsilon_n : n \geq 1\}$ be a Rademacher sequence (i.e., an i.i.d. sequence with $Q(\varepsilon_n = -1) = Q(\varepsilon_n = 1) = 1/2$) independent of $\{Y_n : n \geq 1\}$. Then

$$(i) \quad E_Q \left\| \sum_{i=1}^n (Y_i - EY_i) \right\| \leq 2E_Q \left\| \sum_{i=1}^n \varepsilon_i Y_i \right\|$$

and

$$(ii) \quad E_Q \left\| \sum_{i=1}^n \varepsilon_i Y_i \right\| \leq 2E_Q \left\| \sum_{i=1}^n (Y_i - EY_i) \right\|.$$

Lemma 3.2 (Contraction principle): Let $\{\xi_n : n \geq 1\}$ and $\{\eta_n : n \geq 1\}$ be two i.i.d. sequences of symmetric random variables such that, for each $t > 0$,

$$Q(\{|\eta_n| > t\}) \leq Q(\{|\xi_n| > t\}).$$

Then for each $n \geq 1$ and for all $x_1, x_2, \dots, x_n \in B$,

$$E_Q \left\| \sum_{i=1}^n \eta_i x_i \right\| \leq E_Q \left\| \sum_{i=1}^n \xi_i x_i \right\|.$$

For a proof, see Ledoux and Talagrand (1991).

Lemma 3.3 (Poissonization lemma): For each $n \geq 1$, let $x_1, \dots, x_n \in B$; let $\{Y_{n,j} : 1 \leq j \leq m\}$ be m i.i.d. B -valued random elements with the distribution

$$n^{-1} \sum_{i=1}^n \delta_{x_i}.$$

Let $\{\varepsilon_i : i \geq 1\}$ and $\{\pi_i^s[m/2n] : i \geq 1\}$ respectively be a Rademacher sequence and a sequence of independent symmetrized Poisson random variables with parameters $m/2n$, both independent of $\{Y_{n,j} : 1 \leq j \leq m\}$. Then

$$\frac{1}{\sqrt{2}}(1 - e^{-\frac{m}{n}})E_Q \left\| \sum_{i=1}^n \varepsilon_i x_i \right\| \leq E_Q \left\| \sum_{j=1}^m \varepsilon_j Y_{n,j} \right\| \leq 2E_Q \left\| \sum_{i=1}^n \pi_i^s \left[\frac{m}{2n} \right] x_i \right\| \quad (3.12)$$

(see Proposition 2.2 in Giné and Zinn (1990) and Lemma 4.2 in Praestgaard and Wellner (1993)). Finally, we recall a few well known properties of the symmetrized Poisson random variables. Let $\pi_i[\lambda]$, $i = 1, 2$, be two independent Poisson random variables with mean $\lambda > 0$.

Then $\pi^s[\lambda] := \pi_1[\lambda] - \pi_2[\lambda]$ is a symmetrized Poisson random variable with parameter $\lambda > 0$. It is easily seen that

$$Q(\{|\pi^s[\lambda]| = k\}) = e^{-2\lambda} \left[\sum_{j=0}^{\infty} \frac{\lambda^{2j+k}}{j!(j+k)!} + \sum_{j=k}^{\infty} \frac{\lambda^{2j-k}}{j!(j-k)!} \right], \quad k = 0, 1, 2, \dots \quad (3.13)$$

We will use the following elementary results:

Lemma 3.4 (i) If $0 < \lambda_1 \leq \lambda_2 \leq 1/2$, then, for each $t > 0$,

$$Q(\{|\pi^s[\lambda_1]| > t\}) \leq Q(\{|\pi^s[\lambda_2]| > t\}). \quad (3.14)$$

(ii) For each λ , $0 < \lambda \leq 1/2$, and for any $k = 0, 1, 2, \dots$

$$Q(\{|\pi^s[\lambda]| = k\}) \leq \text{const} \frac{\lambda^k}{k!}. \quad (3.15)$$

Suppose now that $\{\mathcal{F}_n : n \geq 1\}$ is a nonincreasing sequence of classes of functions in $L_2(X; dP)$, i.e. $L_2(X; dP) \supseteq \mathcal{F}_n \supseteq \mathcal{F}_{n+1}$ for all $n \geq 1$. Let ψ be a concave nondecreasing function from $[0, +\infty)$ into $[0, +\infty)$, such that for some $c > 0$ and for all $x \geq 0$, $y \geq 0$

$$\psi(xy) \leq c\psi(x)\psi(y) \quad (3.16)$$

and

$$\sum_{j=1}^{\infty} j\psi\left(\frac{1}{2^j j!}\right) < \infty. \quad (3.17)$$

Lemma 3.5. Suppose that for all $n \geq 1$

$$E^* \left\| \sum_{j=1}^n (\delta_{X_j} - P) \right\|_{\mathcal{F}_n} \leq \psi(n), \quad (3.18)$$

Then for all $m \leq n$

$$E^* \max_{k \leq m} \hat{E} \left\| \sum_{j=1}^k (\delta_{X_{n,j}} - P_n) \right\|_{\mathcal{F}_n} \leq \text{const} \psi(m). \quad (3.19)$$

Proof: To estimate the expectation in (3.19), we need the following sequences of mutually independent random variables:

(a) a Rademacher sequence $\{\varepsilon_j : j \geq 1\}$;

(b) for each $k, 1 \leq k \leq m$, a sequence $\{\pi_j^s[k/2n] : j \geq 1\}$ of independent symmetrized Poisson random variables with parameters $(k/2n)$.

We assume that the Rademacher sequence, symmetrized Poisson random variables, and the bootstrapped sequence $\{\hat{X}_{n,j} : j \geq 1\}$ are all independent of one another. More specifically, we define the Rademacher sequence on a separate probability space $(\Omega_\varepsilon, \Sigma_\varepsilon, \gamma_\varepsilon)$; the

symmetrized Poisson sequences on another probability space $(\Omega_\pi, \Sigma_\pi, \gamma_\pi)$. Thus in what follows, our underlying probability space admits the form:

$$(\Omega, \Sigma, \gamma) \times (\hat{\Omega}, \hat{\Sigma}, \hat{\gamma}) \times (\Omega_\varepsilon, \Sigma_\varepsilon, \gamma_\varepsilon) \times (\Omega_\pi, \Sigma_\pi, \gamma_\pi).$$

Then by the symmetrization lemma, we have a.s. (γ)

$$\max_{k \leq m} \hat{E} \left\| \sum_1^k (\delta_{X_{n,j}} - P_n) \right\|_{\mathcal{F}_n} \leq 2 \max_{k \leq m} \hat{E} E_\varepsilon \left\| \sum_1^k \varepsilon_j (\delta_{X_{n,j}} - P_n) \right\|_{\mathcal{F}_n} \quad (3.20)$$

which, by the Poissonization lemma, is

$$\leq 2 \max_{k \leq m} E_\pi \left\| \sum_{i=1}^n \pi_i^s \left[\frac{k}{2n} \right] \delta_{X_i} \right\|_{\mathcal{F}_n}. \quad (3.21)$$

And now by the contraction principle and Lemma 3.4(i), the expression (3.21) is

$$\leq 2 E_\pi \left\| \sum_{i=1}^n \pi_i^s \left[\frac{m}{2n} \right] \delta_{X_i} \right\|_{\mathcal{F}_n}. \quad (3.22)$$

Since random variables π_i^s are symmetric, we have

$$\begin{aligned} E E_\pi \left\| \sum_{i=1}^n \pi_i^s \left[\frac{m}{2n} \right] \delta_{X_i} \right\|_{\mathcal{F}_n} &\leq E E_\pi E_\varepsilon \left\| \sum_{i=1}^n |\pi_i^s \left[\frac{m}{2n} \right]| \varepsilon_i \delta_{X_i} \right\|_{\mathcal{F}_n} \\ &\leq E E_\pi E_\varepsilon \left\| \sum_{i=1}^n \sum_{j \geq 1} |\pi_i^s \left[\frac{m}{2n} \right]| Z_{ij} \varepsilon_i \delta_{X_i} \right\|_{\mathcal{F}_n}, \end{aligned} \quad (3.23)$$

where $Z_{ij} = 1$ if $\pi_i^s \left[\frac{m}{2n} \right] = \pm j$, and $Z_{ij} = 0$ otherwise. Then (3.23) is

$$= E E_\pi E_\varepsilon \left\| \sum_{j \geq 1} j \sum_{i=1}^n Z_{ij} \varepsilon_i \delta_{X_i} \right\|_{\mathcal{F}_n} = E E_\pi E_\varepsilon \left\| \sum_{j \geq 1} j \sum_{i \in \Delta_n(j)} \varepsilon_i \delta_{X_i} \right\|_{\mathcal{F}_n}, \quad (3.24)$$

where $\Delta_n(j) := \{i : |\pi_i^s \left(\frac{m}{2n} \right)| = j, 1 \leq i \leq n\}$. Thus

$$(3.24) \leq \sum_{j \geq 1} j E_\pi E_\varepsilon E \left\| \sum_{i \in \Delta_n(j)} \varepsilon_i \delta_{X_i} \right\|_{\mathcal{F}_n}. \quad (3.25)$$

Now the symmetrization lemma and condition (3.18) imply that for $K := \text{card}(\Delta_n(j))$

$$\begin{aligned} E_\varepsilon E \left\| \sum_{i \in \Delta_n(j)} \varepsilon_i \delta_{X_i} \right\|_{\mathcal{F}_n} &\leq 2 E \left\| \sum_{i \in \Delta_n(j)} (\delta_{X_i} - P) \right\|_{\mathcal{F}_n} \\ &\leq 2 E \left\| \sum_{i=1}^K (\delta_{X_i} - P) \right\|_{\mathcal{F}_K} \leq 2\psi(K) = 2\psi(\text{card}(\Delta_n(j))). \end{aligned} \quad (3.26)$$

Therefore, by (3.26) and Jensen's inequality, (3.25) is

$$\begin{aligned}
&\leq 2 \sum_{j \geq 1} j E_{\pi} \psi(\text{card}(\Delta_n(j))) \leq 2 \sum_{j \geq 1} j \psi(E_{\pi} \text{card}(\Delta_n(j))) \\
&= 2 \sum_{j \geq 1} j \psi(E_{\pi} \sum_{i=1}^n Z_{ij}) = 2 \sum_{j \geq 1} j \psi(n P(Z_{ij} = 1)) \\
&\leq 2 \sum_{j \geq 1} j \psi(\text{const } n (\frac{m}{2n})^j \frac{1}{j!}), \quad \text{by Lemma 3.4(ii),} \\
&\leq 2 \sum_{j \geq 1} j \psi(\text{const } m (\frac{1}{2})^{j-1} \frac{1}{j!}), \\
&\leq \text{const } \psi(m), \quad \text{by (3.16) and (3.17).}
\end{aligned}$$

Remark: There could be another approach to the proof, based on Lemma 1.2.4 in Giné and Zinn (1986) and Lemma 4.4 in Praestgaard and Wellner (1993). But we preferred to present here a proof, based mostly on the properties of symmetrized Poisson random variables.

The investigation of the classical bootstrap for the general empirical measures was initiated by P. Gaenssler (1987). The following remarkable theorem is due to Giné and Zinn (1990).

Theorem 3.1. The following two conditions are equivalent:

- (a) $\mathcal{F} \in CLT(P)$;
- (b) There exists a Gaussian process $G(f)$, $f \in \mathcal{F}$, defined on the probability space $(\hat{\Omega}, \hat{\Sigma}, \hat{\gamma})$ such that $G \in C_{bu}(\mathcal{F})$ $\hat{\gamma}$ -a.s. and

$$d_{\gamma}(\sqrt{n}(\hat{P}_n - P_n), G) \rightarrow 0,$$

in probability (γ) , as $n \rightarrow \infty$. The process G is a P -Brownian bridge.

Praestgaard and Wellner (1993) studied more general versions of the bootstrap, including the one in which the size of bootstrap sample $m \neq n$.

We adopt the following terminology in the sequel. Given a sequence η_n of random variables defined on the probability space $(\Omega, \Sigma, \gamma) \times (\hat{\Omega}, \hat{\Sigma}, \hat{\gamma})$ and a sequence a_n of positive real numbers, we write

$$\eta_n = o_p(a_n) \tag{3.27}$$

iff, for each $\varepsilon > 0$,

$$\gamma \times \hat{\gamma}(\{|\eta_n| \geq \varepsilon a_n\}) \rightarrow 0,$$

as $n \rightarrow \infty$.

We write

$$\eta_n = O_p(a_n)$$

iff

$$\lim_{c \rightarrow \infty} \limsup_{n \rightarrow \infty} \gamma \times \hat{\gamma}(\{|\eta_n| \geq ca_n\}) = 0.$$

It is easy to check that $\eta_n = O_p(a_n)$ if and only if $\eta_n = o_p(b_n)$, for any sequence b_n of positive numbers such that $a_n = o(b_n)$.

An immediate consequence of Fubini's theorem is that $\eta_n = o_p(a_n)$ iff, for each $\varepsilon > 0$,

$$\hat{\gamma}(\{\hat{\omega} : |\eta_n(\cdot, \hat{\omega})| \geq \varepsilon a_n\}) \rightarrow 0, \quad (3.28)$$

as $n \rightarrow \infty$ in probability (γ).

The following theorem furnishes the main results of this section.

Theorem 3.2. Let $\{a_n\}$ be a sequence of positive real numbers such that $a_n = O(n)$. Suppose that $\mathcal{F} \in CLT(P)$ and let

$$|N_n - n| = o_p(a_n). \quad (3.29)$$

Then

$$\|\hat{P}_{n, N_n} - \hat{P}_n\|_{\mathcal{F}} = o_p\left(\frac{a_n^{1/2}}{n}\right). \quad (3.30)$$

Proof: In view of (3.28), it is enough to show that, for each $\varepsilon > 0$,

$$\hat{\gamma}\left\{\|\hat{P}_{n, N_n} - \hat{P}_n\|_{\mathcal{F}} \geq \frac{\varepsilon a_n^{1/2}}{n}\right\} \rightarrow 0 \quad (3.31)$$

as $n \rightarrow \infty$, in γ .

Let $\hat{P}_{N_n \Delta n}$ denote the empirical measure based on $|N_n - n|$ bootstrapped observations between n and N_n . Then, we have

$$\hat{P}_{n, N_n} = \frac{n}{N_n} \hat{P}_n + \left(1 - \frac{n}{N_n}\right) \hat{P}_{N_n \Delta n}, \quad (3.32)$$

which implies

$$\begin{aligned} \|\hat{P}_{n, N_n} - \hat{P}_n\|_{\mathcal{F}} &= \left|1 - \frac{n}{N_n}\right| \|\hat{P}_{N_n \Delta n} - \hat{P}_n\|_{\mathcal{F}} \\ \|\hat{P}_{n, N_n} - \hat{P}_n\|_{\mathcal{F}} &\leq \frac{|N_n - n|}{n} \left(\frac{n}{N_n}\right) [\|\hat{P}_n - P_n\|_{\mathcal{F}} \\ &\quad + \|\hat{P}_{n \Delta N_n} - P_n\|_{\mathcal{F}}]. \end{aligned} \quad (3.33)$$

From (3.29), it follows that as $n \rightarrow \infty$, $n/N_n = 1 + o_p(1)$, and $|N_n - n|/n = o_p(a_n^{1/2}/n)$. Moreover the Gine-Zinn Theorem (Theorem 3.1) implies that for any sequence $b_n \rightarrow \infty$, we have

$$\|\hat{P}_n - P_n\|_{\mathcal{F}} = o_{\gamma}\left(\frac{b_n}{\sqrt{n}}\right) \quad (3.34)$$

in probability (γ).

Thus, under the condition $a_n = O(n)$, we have

$$\frac{|N_n - n|}{n} \frac{n}{N_n} \|\hat{P}_n - P_n\|_{\mathcal{F}} = o\left(\frac{a_n}{n}\right) \|\hat{P}_n - P_n\|_{\mathcal{F}} = o_p\left(\frac{a_n^{1/2}}{n}\right). \quad (3.35)$$

Therefore it suffices to show that

$$|N_n - n| \|\hat{P}_{N_n \Delta N_n} - P_n\|_{\mathcal{F}} = o_{\gamma}(a_n^{1/2}) \quad (3.36)$$

in probability (γ) .

Now let $m_n = \lfloor a_n \varepsilon^4 \rfloor$. Then

$$\begin{aligned} & \hat{\gamma}(\{|N_n - n| \|\hat{P}_{N_n \Delta N_n} - P_n\|_{\mathcal{F}} \geq \varepsilon a_n^{1/2}\}) \\ &= \hat{\gamma}(\{|N_n - n| \|\hat{P}_{|N_n - n|} - P_n\|_{\mathcal{F}} \geq \varepsilon a_n^{1/2}\}) \\ &= \hat{\gamma}(\{\|\sum_1^{|N_n - n|} (\delta_{X_{n,j}} - P_n)\|_{\mathcal{F}} \geq \varepsilon a_n^{1/2}\}) \\ &\leq \hat{\gamma}(\{|N_n - n| \geq m_n\}) \\ &+ \hat{\gamma}(\{\max_{k \leq m_n} \|\sum_1^k (\delta_{X_{n,j}} - P_n)\|_{\mathcal{F}} \geq \varepsilon a_n^{1/2}\}). \end{aligned} \quad (3.37)$$

By condition (3.29), we have

$$\hat{\gamma}(\{|N_n - n| \geq m_n\}) \rightarrow 0 \quad (3.38)$$

in probability (γ) .

Now to estimate the last term in (3.37), we invoke the following version of Ottaviani's inequality:

$$\begin{aligned} & \hat{\gamma}(\{\max_{k \leq m_n} \|\sum_1^k (\delta_{X_{n,j}} - P_n)\|_{\mathcal{F}} \geq \varepsilon a_n^{1/2}\}) \\ &\leq \frac{\hat{\gamma}(\{\|\sum_1^{m_n} (\delta_{X_{n,j}} - P_n)\|_{\mathcal{F}} \geq \varepsilon a_n^{1/2}\})}{1 - \max_{k \leq m_n} \hat{\gamma}(\{\|\sum_1^k (\delta_{X_{n,j}} - P_n)\|_{\mathcal{F}} \geq \frac{\varepsilon a_n^{1/2}}{2}\})}. \end{aligned} \quad (3.39)$$

Thus it is enough to show that

$$\max_{k \leq m_n} \hat{\gamma}(\{\|\sum_1^k (\delta_{X_{n,j}} - P_n)\|_{\mathcal{F}} \geq \frac{\varepsilon a_n^{1/2}}{2}\}) \rightarrow 0, \quad (3.40)$$

in probability (γ) . Clearly

$$\begin{aligned} & \max_{k \leq m_n} \hat{\gamma}(\{\|\sum_1^k (\delta_{X_{n,j}} - P_n)\|_{\mathcal{F}} \geq \frac{\varepsilon a_n^{1/2}}{2}\}) \\ &\leq \frac{2}{\varepsilon} a_n^{-1/2} \max_{k \leq m_n} \hat{E} \|\sum_1^k (\delta_{X_{n,j}} - P_n)\|_{\mathcal{F}}. \end{aligned} \quad (3.41)$$

To estimate this last expectation, we apply Lemma 3.5 to the classes $\mathcal{F}_n := \mathcal{F}$ and $\psi(x) := \text{const } x^{1/2}$. Condition (3.18) holds by (3.11), and it follows from (3.41) and (3.19) that

$$E \max_{k \leq m_n} \hat{\gamma}(\{\|\sum_1^k (\delta_{X_{n,j}} - P_n)\|_{\mathcal{F}} \geq \frac{\varepsilon a_n^{1/2}}{2}\}) \leq \text{const } \varepsilon.$$

Since ε can be chosen arbitrarily small, (3.40) follows.

This completes the proof of Theorem 3.2.

Corollary 3.2.1: Let $\{a_n\}$ be a sequence of positive numbers such that $a_n = o(n)$. Suppose that $\mathcal{F} \in CLT(P)$ and $|N_n - n| = O_p(a_n)$. Then

$$\|\hat{P}_{n,N_n} - \hat{P}_n\|_{\mathcal{F}} = O_p\left(\frac{a_n^{1/2}}{n}\right).$$

To prove the corollary, it is enough to apply Theorem 3.2 to any sequence $\{b_n\}$ such that $a_n = o(b_n)$.

Corollary 3.2.2: Suppose that $\mathcal{F} \in CLT(P)$ and $|N_n - n| = o_p(n)$. Then

$$\|\hat{P}_{n,N_n} - \hat{P}_n\|_{\mathcal{F}} = o_p(n^{-1/2}).$$

Corollary 3.2.3: Under the conditions of Corollary 3.2.2,

$$\|N_n^{1/2}(\hat{P}_{n,N_n} - P_n) - n^{1/2}(\hat{P}_n - P_n)\|_{\mathcal{F}} = o_p(1).$$

Moreover, under the conditions of Theorem 3.2,

$$\|N_n^{1/2}(\hat{P}_{n,N_n} - P_n) - n^{1/2}(\hat{P}_n - P_n)\|_{\mathcal{F}} = o_p\left(\sqrt{\frac{a_n}{n}}\right),$$

and under the conditions of Corollary 3.2.1

$$\|N_n^{1/2}(\hat{P}_{n,N_n} - P_n) - n^{1/2}(\hat{P}_n - P_n)\|_{\mathcal{F}} = O_p\left(\sqrt{\frac{a_n}{n}}\right),$$

Indeed,

$$\begin{aligned} & \|N_n^{1/2}(\hat{P}_{n,N_n} - P_n) - n^{1/2}(\hat{P}_n - P_n)\|_{\mathcal{F}} \\ & \leq |N_n^{1/2} - n^{1/2}|(\|\hat{P}_{n,N_n} - P_n\|_{\mathcal{F}} + \|\hat{P}_n - P_n\|_{\mathcal{F}}) + n^{1/2}\|\hat{P}_{n,N_n} - \hat{P}_n\|_{\mathcal{F}}. \end{aligned}$$

Under the conditions of Corollary 3.2.2,

$$|N_n^{1/2} - n^{1/2}| = n^{1/2}|\left(\frac{N_n}{n}\right)^{1/2} - 1| = o_p(n^{1/2}).$$

Under the conditions of Theorem 3.2 we have

$$|N_n^{1/2} - n^{1/2}| = o_p(a_n n^{-1}) = o_p\left(\sqrt{\frac{a_n}{n}}\right).$$

Theorem 3.1 implies that $\{n^{1/2}\|\hat{P}_n - P_n\|_{\mathcal{F}}\}$ is stochastically bounded, and the result follows from Corollary 3.2.2 and Theorem 3.2. The last case is quite similar.

Corollary 3.2.4: Under conditions of Corollary 3.2.2,

$$d_{\gamma \times \gamma}(\sqrt{N_n}(\hat{P}_{n,N_n} - P_n), \sqrt{n}(\hat{P}_n - P_n)) \rightarrow 0,$$

and there exists a P -Brownian bridge $G_P(f)$, $f \in \mathcal{F}$ defined on the probability space $(\hat{\Omega}, \hat{\Sigma}, \hat{\gamma})$ such that $G_P \in C_{bu}(\mathcal{F})$ $\hat{\gamma}$ -a.s. and

$$d_{\gamma}(\sqrt{N_n}(\hat{P}_{n,N_n} - P_n), G_P) \rightarrow 0$$

in probability (γ) .

4. MATHEMATICAL JUSTIFICATION: CONVERGENCE RATES FOR THE SEQUENTIAL BOOTSTRAP

In what follows we apply the results of Section 3 to the empirical measures based on sequential resampling bootstrap. We start with the following theorem summarizing the properties of the empirical measures in this case.

Theorem 4.1. Suppose that $\mathcal{F} \in CLT(P)$ and let \hat{P}_{n,N_n} be the empirical measure based on a sequential bootstrap sample. Then

$$\|\hat{P}_{n,N_n} - \hat{P}_n\|_{\mathcal{F}} = O_p(n^{-3/4})$$

and

$$\|N_n^{1/2}(\hat{P}_{N_n,n} - P_n) - n^{1/2}(\hat{P}_n - P_n)\|_{\mathcal{F}} = O_p(n^{-1/4}).$$

Indeed, in this case we have, by (2.9), $|N_n - n| = O_p(n^{1/2})$ and Corollaries 3.2.1 and 3.2.3 imply the result.

In particular, we have (see Corollary 3.2.4)

$$d_{\gamma \times \gamma}(\sqrt{N_n}(\hat{P}_{n,N_n} - P_n), \sqrt{n}(\hat{P}_n - P_n)) \rightarrow 0$$

and

$$d_{\gamma}(\sqrt{N_n}(\hat{P}_{n,N_n} - P_n), G_P) \rightarrow 0,$$

in probability (γ) , where $G_P(f)$, $f \in \mathcal{F}$ is a P -Brownian bridge, defined on the probability space $(\hat{\Omega}, \hat{\Sigma}, \hat{\gamma})$ and such that $G_P \in C_{bu}(\mathcal{F})$ $\hat{\gamma}$ -a.s..

For simplicity, we consider in what follows only the case $X = R^1$. In this case, it is natural to take $\mathcal{F} := \{I_{(-\infty, t]} : t \in R^1\}$, so that general empirical measures considered above turn out to be classical empirical distribution functions. Denote

$$F(t) := P((-\infty, t]) = \int_{R^1} I_{(-\infty, t]} dP,$$

$$F_n(t) := P_n((-\infty, t]) = \int_{R^1} I_{(-\infty, t]} dP_n$$

and

$$\hat{F}_{n,N}(t) := \hat{P}_{n,N}((-\infty, t]) = \int_{R^1} I_{(-\infty, t]} d\hat{P}_{n,N}.$$

We also use the abbreviation $\hat{F}_n := \hat{F}_{n,n}$.

In what follows $\|\cdot\|_\infty$ denotes the sup-norm of a function from an interval $J \subset R^1$ into R^1 (in particular, there could be $J = R^1$).

Since, by the Kolmogorov-Donsker theorem, we have $\mathcal{F} \in CLT(P)$ for all Borel probability measures P on R^1 (where $\mathcal{F} = \{I_{(-\infty, t]} : t \in R^1\}$), we have the following statement.

Theorem 4.2. For any distribution function F on R^1 ,

$$\|\hat{F}_{n,N_n} - \hat{F}_n\|_\infty = O_p(n^{-3/4})$$

and

$$\|N_n^{1/2}(\hat{F}_{n,N_n} - F_n) - n^{1/2}(\hat{F}_n - F_n)\|_\infty = O_p(n^{-1/4}).$$

In the case of uniform distribution F on $[0, 1]$, we get as a trivial corollary that the sequence of stochastic processes

$$\{N_n^{1/2}(\hat{F}_{n,N_n}(t) - F_n(t)) : t \in [0, 1]\}_{n \geq 1}$$

converges weakly (say, in the space $\ell^\infty([0, 1])$ or $D[0, 1]$) to the standard Brownian bridge process

$$B(t) := w(t) - tw(1), \quad t \in [0, 1],$$

w being the standard Wiener process. More generally, if F is a continuous distribution function in R^1 , then the limit is the process $(B \circ F)(t) = B(F(t))$, $t \in R^1$. These facts easily imply justification of the sequential resampling bootstrap for a variety of statistics θ_n , which can be represented as $\theta_n = T(F_n)$ with a compactly (Hadamard) differentiable functional T . More precisely, T is a functional (or, more generally, an operator with values in a linear normed space) $G \mapsto T(G)$, defined on a set \mathcal{G} of distribution functions G . T is supposed to be compactly differentiable at F tangentially to the space of all uniformly bounded and uniformly continuous functions (see, e.g., Gill (1989)). For such statistics, we have

$$T(\hat{F}_{n,N_n}) - T(\hat{F}_n) = o_p(n^{-1/2}),$$

proving the first order asymptotic correctness of the sequential resampling approach. These observations can be applied, for instance, to the operator $G \mapsto G^{-1}$, defined by

$$G^{-1}(t) := \inf\{x \in R^1 : G(x) \geq t\},$$

and taking a distribution function to its quantile function (see, e.g., Fernholz (1983) or Gill (1989) for compact differentiability of such operators; see, also, Koltchinskii (1995) for the extensions to the multivariate case). Specifically, if F is continuously differentiable at a point $x \in R^1$ with $F'(x) > 0$, then

$$|\hat{F}_{n,N_n}^{-1}(t) - \hat{F}_n^{-1}(t)| = o_p(n^{-1/2}), \quad (4.1)$$

where $t = F(x)$. In fact, we obtain below a bound for quantiles, sharper than (4.1) (see Theorem 4.4).

Another interesting application of Theorem 4.1 is related to a recent work of Dudley (1992). Namely, we show that in Theorem 4.2 the sup-norm can be replaced by a stronger one, the so called p -variation norm. For a function f from an interval $J \subset R^1$ into R^1 and for $p > 0$, define the p -variation of f as

$$v_p(f) := \sup \left\{ \sum_{i=1}^m |f(x_i) - f(x_{i-1})|^p : x_0 < x_1 < \dots < x_m \in J, x_0 \in J, m = 1, 2, \dots \right\}.$$

Let

$$W_p := W_p(J) := \{f : v_p(f) < +\infty\}.$$

We introduce the following norm on W_p for $p \geq 1$ (see Dudley (1992)):

$$\|f\|_{[p]} := \|f\|_{\infty} + v_p^{1/p}(f).$$

Theorem 4.3. For any distribution function F on R^1 and for all $p > 2$

$$\|\hat{F}_{n,N_n} - \hat{F}_n\|_{[p]} = O_p(n^{-3/4})$$

and

$$\|N_n^{1/2}(\hat{F}_{n,N_n} - F_n) - n^{1/2}(\hat{F}_n - F_n)\|_{[p]} = O_p(n^{-1/4}).$$

Proof: Assume that $J = R^1$, take $q < 2$ and consider

$$\mathcal{F}_q := \{f : \|f\|_{[q]} \leq 1\}.$$

By Theorem 2.1 of Dudley (1992), $\mathcal{F}_q \in CLT(P)$ for any Borel probability measure P on R^1 . On the other hand, L.C. Young's duality inequalities for p -variation norms (see Dudley (1992), Proposition 3.7) imply that for any $p > 1$ and $q > 1$ with $p^{-1} + q^{-1} = 1$ we have

$$\|\hat{F}_{n,N_n} - \hat{F}_n\|_{[p]} \leq \text{const} \|\hat{P}_{n,N_n} - \hat{P}_n\|_{\mathcal{F}_q},$$

and Theorem 4.1 implies the result.

Finally, we improve bound (4.1) for empirical quantiles.

Theorem 4.4. Suppose that F is an absolutely continuous distribution function in R^1 . Let

$$A := \sup\{x : F(x) = 0\}, B := \inf\{x : F(x) = 1\}.$$

Suppose the density of F is a strictly positive continuous function in (A, B) . Then for all $t \in (0, 1)$

$$|\hat{F}_{n,N_n}^{-1}(t) - \hat{F}_n^{-1}(t)| = O_p(n^{-3/4}). \quad (4.2)$$

For any closed interval $J \subset (0, 1)$

$$\sup_{t \in J} |\hat{F}_{n,N_n}^{-1}(t) - \hat{F}_n^{-1}(t)| = O_p(n^{-3/4} \sqrt{\log n}). \quad (4.3)$$

If, moreover, $A > -\infty$, $B < +\infty$ and the density of F is a strictly positive continuous function on $[A, B]$, then (4.3) holds with $J = [0, 1]$.

In order to prove this theorem we need several lemmas. The first one gives an inequality for quantiles. It requires some notations. Given $h : [0, 1] \mapsto R^1$, denote

$$\omega_h(s; \delta) := \sup\{|h(u) - h(v)| : |u - s| \leq \delta, |v - s| \leq \delta, u, v \in [0, 1]\},$$

$$\omega_h(\delta) := \sup_{s \in [0, 1]} \omega_h(s; \delta), \quad \delta > 0.$$

Given a distribution function G , we set

$$\delta(G) := \sup_{t \in (0, 1)} |G(G^{-1}(t)) - t|.$$

Lemma 4.1. Suppose that F is a uniform distribution function on $[0, 1]$. Then, for all distribution functions G and H on $[0, 1]$ and for all $t \in [0, 1]$, we have

$$\begin{aligned} & |G^{-1}(t) - H^{-1}(t)| \\ & \leq \omega_{G-F}(t; \|G - F\|_\infty + \|H - F\|_\infty + \delta(G) + \delta(H)) \\ & \quad + \|H - G\|_\infty + \delta(G) + \delta(H). \end{aligned} \quad (4.4)$$

Proof: First, note that for all $t \in [0, 1]$

$$|G^{-1}(t) - t| \leq |G^{-1}(t) - G(G^{-1}(t))| + |G(G^{-1}(t)) - t| \leq \|G - F\|_\infty + \delta(G).$$

Similarly,

$$|H^{-1}(t) - F^{-1}(t)| \leq \|H - F\|_\infty + \delta(H).$$

It follows that

$$|G^{-1}(t) - H^{-1}(t)| \leq \|G - F\|_\infty + \|H - F\|_\infty + \delta(G) + \delta(H). \quad (4.5)$$

On the other hand,

$$|G^{-1}(t) - H^{-1}(t)| \leq |(G - F)(G^{-1}(t)) - (H - F)(H^{-1}(t))| + |G(G^{-1}(t)) - H(H^{-1}(t))|. \quad (4.6)$$

Since, by (4.5),

$$\begin{aligned} & |(G - F)(G^{-1}(t)) - (H - F)(H^{-1}(t))| \\ & \leq |(G - F)(G^{-1}(t)) - (G - F)(H^{-1}(t))| + |(G - F)(H^{-1}(t)) - (H - F)(H^{-1}(t))| \\ & \leq \omega(t; \|G - F\|_\infty + \|H - F\|_\infty + \delta(G) + \delta(F)) + \|G - H\|_\infty \end{aligned}$$

and

$$|G(G^{-1}(t)) - H(H^{-1}(t))| \leq |G(G^{-1}(t)) - t| + |t - H(H^{-1}(t))| \leq \delta(G) + \delta(H),$$

(4.6) implies (4.4), which completes the proof.

Next we need the following bounds for the continuity modulus of the standard empirical process

$$\zeta_n(t) := n^{1/2}(F_n(t) - t), \quad t \in [0, 1],$$

based on a sample from the uniform distribution on $[0, 1]$. They seem to be known, although we have not found the exact reference. See Stute (1982) and Shorack and Wellner (1986) for some other results on oscillatory behaviour of empirical processes. We sketch the proof here for completeness.

Lemma 4.2. For any $C > 0$ and for all $t \in [0, 1]$

$$E\omega_{\zeta_n}(t; Cn^{-1/2}) = O(n^{-1/4}). \quad (4.7)$$

Moreover,

$$E\omega_{\zeta_n}(Cn^{-1/2}) = O(n^{-1/4} \sqrt{\log n}). \quad (4.8)$$

Proof: By the Komlos-Major-Tusnady theorem (see, e.g. Shorack and Wellner(1986)), there exists a sequence $\{B_n\}_{n \geq 1}$ of Brownian bridges on $[0, 1]$ and constants $D > 0, K > 0, \theta > 0$ such that

$$\gamma\{\Delta_n \geq n^{-1/2}(x + D \log n)\} \leq K e^{-\theta x}, \quad x > 0,$$

where $\Delta_n := \|\zeta_n - B_n\|_\infty$. It follows that

$$E \max(n^{1/2} \Delta_n - D \log n, 0) \leq \frac{K}{\theta},$$

and since

$$n^{1/2} \Delta_n \leq D \log n + \max(n^{1/2} \Delta_n - D \log n, 0),$$

we have

$$E \Delta_n = O(n^{-1/2} \log n).$$

Note that

$$\omega_{\zeta_n}(t; Cn^{-1/2}) \leq \omega_{B_n}(t; Cn^{-1/2}) + 2\Delta_n.$$

and

$$\omega_{\zeta_n}(Cn^{-1/2}) \leq \omega_{B_n}(Cn^{-1/2}) + 2\Delta_n.$$

Therefore it's enough to show that

$$E\omega_B(t; Cn^{-1/2}) = O(n^{-1/4})$$

and

$$E\omega_B(Cn^{-1/2}) = O(n^{-1/4} \sqrt{\log n}),$$

which follow from the representation $B(t) = w(t) - tw(1)$ and well known properties of the Wiener process.

Proof of Theorem 4.4: Without loss of generality, we can and do assume that F is uniform on $[0, 1]$ (in the general case, the transformation $X_i \mapsto F(X_i)$ should be used to reduce the problem to the uniform one). Then we use Lemma 4.1 to get the following bound:

$$\begin{aligned} & |\hat{F}_{n, N_n}^{-1}(t) - \hat{F}_n^{-1}(t)| \\ & \leq \omega_{\hat{F}_n - F}(t; \|\hat{F}_{n, N_n} - F\|_\infty + \|\hat{F}_n - F\|_\infty + \delta(\hat{F}_{n, N_n}) + \delta(\hat{F}_n)) \\ & \quad + \|\hat{F}_{n, N_n} - \hat{F}_n\|_\infty + \delta(\hat{F}_{n, N_n}) + \delta(\hat{F}_n). \end{aligned} \quad (4.9)$$

Note that, by usual limit theorems for empirical processes and their bootstrapped versions,

$$\|\hat{F}_n - F\|_\infty \leq \|\hat{F}_n - F_n\|_\infty + \|F_n - F\|_\infty = O_p(n^{-1/2}). \quad (4.10)$$

By Theorem 4.2,

$$\|\hat{F}_{n, N_n} - \hat{F}_n\|_\infty = O_p(n^{-3/4}). \quad (4.11)$$

It follows that we also have

$$\|\hat{F}_{n, N_n} - F\|_\infty = O_p(n^{-1/2}). \quad (4.12)$$

Now we are going to show that

$$\delta(\hat{F}_n) = O_p(n^{-1} \log n) \quad (4.13)$$

and

$$\delta(\hat{F}_{n, N_n}) = O_p(n^{-1} \log n). \quad (4.14)$$

Indeed, $\delta(\hat{F}_n)$ is less than or equal to the largest jump of \hat{F}_n , which is equal to

$$\frac{\max_{1 \leq j \leq n} \nu_j^{(n)}}{n},$$

where

$$\nu_j^{(n)} := \sum_{k=1}^n I_{\{X_{n,k} = X_j\}}, \quad j = 1, \dots, n.$$

We have

$$\hat{E} \exp\{\nu_j^{(n)}\} = \prod_{k=1}^n \hat{E} \exp\{I_{\{X_{n,k} = X_j\}}\} = \left(1 + \frac{e-1}{n}\right)^n \leq e^{e-1}.$$

Therefore, for $C > 1$

$$\begin{aligned} & \hat{\gamma}\left\{\max_{1 \leq j \leq n} \nu_j^{(n)} \geq C \log n\right\} \\ & \leq \exp\{-C \log n\} n \hat{E} \exp\{\nu_j^{(n)}\} \leq e^{e-1} n^{-C+1} = o(1), \end{aligned}$$

γ -a.e. and (4.13) follows.

The proof of (4.14) is quite similar. In this case we set

$$\nu_j^{(n)} := \sum_{k=1}^{N_n} I_{\{X_{n,k}=X_j\}}, \quad j = 1, \dots, n$$

and get

$$\delta(\hat{F}_{n,N_n}) \leq \frac{\max_{1 \leq j \leq n} \nu_j^{(n)}}{N_n}. \quad (4.15)$$

We have

$$\hat{\gamma}\left\{\max_{1 \leq j \leq n} \nu_j^{(n)} \geq C \log n\right\} \leq \hat{\gamma}\left\{\max_{1 \leq j \leq n} \tilde{\nu}_j^{(n)} \geq C \log n\right\} + \hat{\gamma}\{N_n > 2n\}, \quad (4.16)$$

where

$$\tilde{\nu}_j^{(n)} := \sum_{k=1}^{2n} I_{\{X_{n,k}=X_j\}}, \quad j = 1, \dots, n.$$

Since

$$\hat{E} \exp\{\tilde{\nu}_j^{(n)}\} \leq e^{2(e-1)},$$

we easily obtain

$$\hat{\gamma}\left\{\max_{1 \leq j \leq n} \tilde{\nu}_j^{(n)} \geq C \log n\right\} = o(1), \quad (4.17)$$

and (4.14) follows from (4.15)-(4.17) and (2.9).

By (4.9)-(4.14), in order to prove (4.2) it remains to show that for all $t \in (0, 1)$

$$\omega_{F_n-F}(t; Cn^{-1/2}) = O_p(n^{-3/4}) \quad (4.18)$$

and in order to prove (4.3) we have to show that

$$\omega_{F_n-F}(Cn^{-1/2}) = O_p(n^{-3/4} \sqrt{\log n}). \quad (4.19)$$

We show here only (4.19), since the proof of (4.18) is quite similar. Note that

$$\omega_{F_n-F}(Cn^{-1/2}) = \|\hat{P}_n - P\|_{\mathcal{F}_n},$$

where

$$\mathcal{F}_n := \{I_{[0,t]} - I_{[0,s]} : |t-s| \leq 2Cn^{-1/2}, \quad s, t \in [0, 1]\}.$$

It follows from Lemma 4.2 that

$$E \left\| \sum_{j=1}^n (\delta_{X_j} - P) \right\|_{\mathcal{F}_n} = n E \|P_n - P\|_{\mathcal{F}_n} = n^{1/2} E \omega_{\zeta_n}(Cn^{-1/2}) = O(n^{1/4} \sqrt{\log n}).$$

Now we use Lemma 3.5 with $m = n$ and with function

$$\psi(x) := \text{const } x^{1/4} \sqrt{\log(x + e^2)}$$

(which satisfies all conditions required in Lemma 3.5), to get

$$nE\hat{E} \|\hat{P}_n - P_n\|_{\mathcal{F}_n} = \left\| \sum_{j=1}^n (\delta_{\hat{X}_{n,j}} - P_n) \right\|_{\mathcal{F}_n} \leq \text{const } \psi(n),$$

which implies

$$E\hat{E}\omega_{\hat{F}_n - F_n}(Cn^{-1/2}) = E\hat{E} \|\hat{P}_n - P_n\|_{\mathcal{F}_n} = O(n^{-3/4} \sqrt{\log n}).$$

Combining this bound with

$$E\hat{E}\omega_{F_n - F}(Cn^{-1/2}) = O(n^{-3/4} \sqrt{\log n})$$

(which follows from (4.8)), we get

$$E\hat{E} \omega_{F_n - F}(Cn^{-1/2}) \leq E\hat{E} \omega_{F_n - F_n}(Cn^{-1/2}) + E\hat{E} \omega_{F_n - F}(Cn^{-1/2}) = O(n^{-3/4} \sqrt{\log n}).$$

This implies (4.19), completing the proof.

It worth noting that the rate $n^{-3/4}$ (up to a logarithmic factor) in Theorem 4.4 is exactly the same as in the remarkable Bahadur-Kiefer representation for empirical quantiles (see, e.g., Shorack and Wellner (1986)).

5. CONCLUDING REMARKS

Having seen in the preceding investigation that the “distance” between the sequential bootstrap and the common bootstrap of size n is at most of order $n^{-3/4}$ (up to a logarithmic factor), it is now natural to ask about the second-order correctness of the sequential bootstrap as compared to the common bootstrap. Our preliminary investigation along the lines of the Hall-Mammen paper (1994) seems to indicate that the second-order correctness of the sequential bootstrap does indeed go through. Further work in this direction is in progress and will be published later.

Acknowledgement. We wish to thank E. Giné for his helpful comments on the proof of Theorem 3.2. Research work of this paper is sponsored by the Army Research Office under Grant DAAHO4-96-1-0082. The work of Pathak and Koltchinskii is partially supported by a Sandia National Laboratories Grant AR-2011.

REFERENCES

- Bickel, P.J. and Freedman, D.A. (1981). Some asymptotic theory for the bootstrap. *Ann. Statist.* 9, 1196-1217.
- Dudley, R.M. (1984). A course on empirical processes, *Lecture Notes in Math.* 1097, 1-142, Springer.
- Dudley, R.M. (1992). Fréchet differentiability, p -variation and uniform Donsker classes. *Ann. Prob.* 20, 1968-1982.
- Efron, B. (1979). Bootstrap methods: Another look at the jackknife. *Ann. Statist.* 7, 1-26.
- Efron, B. (1982). *The Jackknife, the Bootstrap, and Other Resampling Plans*. SIAM, Philadelphia.
- Efron, B. (1983). Estimating the error rate of a prediction rule: Improvement on cross-validation. *J. Am. Statist. Assoc.* 78, 316-331.
- Fernholz, L.T. (1983). Von Mises calculus for statistical functionals, *Lecture Notes in Statistics* 19, Springer-Verlag, New York.
- Gaenssler, P. (1987). Bootstrapping empirical measures indexed by Vapnik-Červonenkis classes of sets. In *Probability Theory and Mathematical Statistics* (Yu. V. Prohorov, V.A. Statulevichius, V. Sazonov et al., Eds) 1, 467-481. VNU Science Press, Utrecht.
- Gill, R. (1989). Non- and Semi-parametric Maximum Likelihood Estimators and the von Mises Method (Part 1). *Scand. J. Statist.* 16, 97-128.
- Giné, E. and Zinn, J. (1984). Some limit theorems for empirical processes. *Ann. Prob.* 12, 929-989.
- Giné, E. and Zinn, J. (1986) Lectures on the central limit theorem for empirical processes. *Lecture Notes in Mathematics* 1221, 50-113. Springer.
- Giné, E. and Zinn, J. (1990) Bootstrapping general empirical measures. *Ann. Prob.* 18, 851-869.
- Hall, Peter and Mammen, Enno (1994) On general resampling algorithms and their performance in distribution estimation. *Ann. Statist.* 24, 2011-2030.
- Koltchinskii, V. (1995) Differentiability of inverse operators and limit theorems for inverse functions. (Preprint).
- Ledoux, M. and Talagrand, M. (1991) *Probability in Banach Spaces. Isoperimetry and Processes*. Springer-Verlag.
- Mallows, C.L. (1972). A note on asymptotic joint normality. *Ann. Math. Stat.* 43, 508-515.
- Mitra, S.K. and Pathak, P.K. (1984). The nature of simple random sampling. *Ann. Statist.* 12, 1536-1542.
- Pathak, P.K. (1964). Sufficiency in sampling theory. *Ann. Math. Statist.* 35, 795-808.
- Pathak, P.K. (1964). On inverse sampling with unequal probabilities. *Biometrika* 51, 185-193.
- Pollard, D. (1984). *Convergence of Stochastic Processes*. Springer, Berlin.
- Praetgaard, J. and Wellner, J. (1993) Exchangeably weighted bootstraps of the general empirical process. *Ann. Prob.* 21, 2053-2086.

- Rachev, S.T. (1984) The Monge-Kantorovich mass transference problem and its stochastic applications. *Theory Probab. and Applic.* 29, 647-676.
- Shorack, G.R. and Wellner, J.A. (1986) Empirical Processes with Applications, Wiley, New York.
- Stute, W. (1982) The oscillation behaviour of empirical processes. *Ann. Prob.* 10, 86-107.